Описание алгоритма работы модуля обработки файлов с информацией о телефонных соединениях

Модуль предназначен для обработки файлов с синтетическими (сгенерированными искусственно) записями о телефонных соединениях (вызовах) и предоставления статистической информации о количестве обработанных соединений и суммарной длительности вызовов в секундах по каждому направлению, определяемому соответствующим набором префиксов. «Синтетические данные (CSV)» представляют собой текстовые CDR (Call Detail Record) файлы, в которых в качестве разделителя полей используется запятая.

Описание значимых полей CDR файла:

|  |  |  |
| --- | --- | --- |
| **№** | **Name** | **Description** |
| 6 | MSISDN | Номер вызывающего абонента |
| 7 | DIALED | Номер вызываемого абонента |
| 8 | START\_TIME | Время начала телефонного вызова |
| 9 | DURATION | Длительность соединения в секундах |
| 31 | CIRCUIT\_IN | Входящая транковая группа |
| 32 | CIRCUIT\_OUT | Исходящая транковая группа |

Модуль последовательно обрабатывает CDR файлы из каталога «Синтетические данные (CSV)», анализируя их построчно. Каждая строка в CDR файле является записью о событии - телефонном соединении. При разборе записи о событии, для полей MSISDN и DIALED, которые являются телефонными номерами вызывающей и принимающей стороны, находится наиболее полное соответствие префикса номера в файле PREFIXES.TXT. Под префиксом понимаются первые цифры телефонного номера, указывающие на принадлежность всех номеров, начинающихся на данный префикс, той или иной префиксной зоне.

Файл PREFIXES.TXT находится в каталоге «Префиксы телефонных номеров (CSV)» и представляет собой текстовый файл, в котором указаны префиксные зоны и принадлежащие этим зонам префиксы различной длины. Префиксная зона отделяется от префикса запятой.

Например, записи

Russia4,74

Russia6,74061954

обозначают, что

префикс **74** относится к префиксной зоне **Russia4**, а

префикс **74061954** относится к префиксной зоне **Russia6**.

Префикс номера можно рассматривать как маску, применяемую к номерам MSISDN и DIALED для определения их принадлежности той или иной префиксной зоне. Если взять в качестве примера MSISDN=74061954000, то видно, что этот номер соответствует как префиксу **74**, так и префиксу **74061954**, так как оба этих префикса совпадают с начальными цифрами номера. Но при анализе предпочтение отдаётся префиксу наибольшей длины – тому, который описывает принадлежность телефонного номера префиксной зоне с наибольшей точностью, то есть в данном случае делается вывод о том, что номер MSISDN=74061954000 относится к префиксной зоне **Russia6**, т.к. префикс **74061954** длиннее префикса **74**.

Определение принадлежности номеров MSISDN и DIALED префиксным зонам по префиксам делается для каждого события в CDR файлах, и в результате для каждого события формируются значения:

1. Наибольший по длине префикс, соответствующий началу строки MSISDN – PREFIX1
2. Префиксная зона, к которой относится PREFIX1 и соответственно номер MSISDN – ZONE1
3. Наибольший по длине префикс, соответствующий началу строки DIALED – PREFIX2
4. Префиксная зона, к которой относится PREFIX2 и соответственно номер DIALED – ZONE2

На выходе модуля получается набор CDR файлов с теми же именами и с таким же количеством записей, при этом получаемые в результате работы модуля записи в файлах обогащаются информацией о префиксных зонах, к которым относятся номера MSISDN и DIALED. Найденные значения префиксов и префиксных зон прописываются в полях:

№10 (префикс PREFIX1, соответствующий номеру MSISDN)

№11 (префиксная зона ZONE1, к которой относится номер MSISDN)

№12 (префикс PREFIX2, соответствующий номеру DIALED)

№13 (префиксная зона ZONE2, к которой относится номер DIALED)

Например, запись

07,4,,,,74061954988,336300254202730,20230401054315,30,,,,,,,,,,,,,,,,,,,,,,{I}INT\_TRUNK\_01,

TELECOM\_02,,,,,100,

будет преобразована в

07,4,,,,74061954988,336300254202730,20230401054315,30,74061954,Russia6,33,Europe2,,,,,,,,,,,,,,,,,,{I}INT\_TRUNK\_01,TELECOM\_02,,,,,100,

т.к. в данном случае номер MSISDN = 74061954988 относится к префиксной зоне Russia6 по соответствию префиксу 74061954,

а номер DIALED = 336300254202730 относится к префиксной зоне Europe2 по соответствию префиксу 33.

Кроме CDR файлов, обогащённых значениями префиксных зон, на выходе модуля после обработки всех записей генерируется файл VOLUMES.TXT со статистикой о количестве соединений TotalCount и количестве секунд TotalDuration, прошедших по каждой найденной паре префиксных зон, то есть агрегированные данные с указанием количества и общей длительности всех соединений.

Формат файла VOLUMES.TXT:

ZONE1, ZONE2,TotalCount,TotalDuration

Пример выходных данных в VOLUMES.TXT:

Russia6,Europe2,4,120

Как уже указывалось, длительность каждого телефонного соединения прописана в поле №9 DURATION. Для приведённого выше примере: DURATION=30.

Если для номера MSISDN или DIALED не был найден ни один соответствующий префикс в файле PREFIXES.TXT, то по умолчанию принимается значение префиксной зоны равное Unknown - эта префиксная зона без префикса указана последней в файле PREFIXES.TXT.